What is Claimed is:

1. A method of generating an index of a class of objects appearing in a collection of images for use in searching or browsing for particular members of said class, comprising:

processing said collection of images to extract therefrom features characteristic of said class of objects; and

grouping said images in groups according to extracted features helpful in identifying individual members of said class of objects.

2. The method according to claim 1, wherein said collection of images represents a sequence of video frames.

3. The method according to claim 2, wherein said grouping of images in groups, according to extracted features helpful in identifying individual members of a class, produces a track of continuous frames for each individual member of said class of objects to be identified.

4. The method according to claim 1, wherein said groups of images are stored in a database store for use in searching or browsing for individual members of the class.

5. The method according to claim 1, wherein said class of objects is human faces.

6. The method according to claim 5, wherein said collection of images represents a sequence of video frames, and said grouping includes forming face tracks of contiguous frames, each track being identified by the starting and ending frames and containing face regions.

7. The method according to claim 6, wherein said sequence of video frames is processed to also include audio data associated with said face tracks.

8. The method according to claim 7, wherein said sequence of video frames is processed to include audio data associated with a said face track by:

generating a face index from the sequence of video frames;

generating a transcription of the audio data; and

aligning said transcription of the audio data with said face index.

9. The method according to claim 8, wherein said transcription of the audio data is generated by closed-caption decoding.

10. The method according to claim 8, wherein said transcription of the audio data is generated by speech recognition.

11. The method according to claim 8, wherein said aligning is effected by:

identifying start and end points of speech segments in the sound data;

extracting face start and end point from the face index; and

outputting all speech segments that have non-zero temporal intersection with the respective face track.

12. The method according to claim 7, wherein said sequence of video frames is also processed to label face tracks as talking or non-talking tracks by:

tracking said face regions in the video frames;

detecting regions having mouth motion; and

estimating from said detected regions those having talking mouth motion vs. non-talking mouth motion.

13. The method according to claim 12, wherein regions estimated to have talking mouth motion are enabled for attaching speech to a speaker in a face track.

14. The method according to claim 7, wherein said sequence of video frames is processed by:

extracting all audio data from video frame regions of a particular face;

fitting a model based on said extracted audio data; and

associating said model with the respective face track.

15. The method according to claim 14, wherein said audio data are speech utterance.

16. The method according to claim 6, wherein said grouping further includes merging tracks containing similar faces from a plurality of said face tracks.

17. The method according to claim 6, wherein said sequence of video frames is processed to include annotations associated with said face track.

18. The method according to claim 6, wherein said sequence of video frames is processed to include face characteristic views associated with said face tracks.

19. The method according to claim 6, wherein said face regions include eye, nose and mouth templates characteristic of individual faces.

20. The method according to claim 6, wherein said face regions include image coordinates of geometric face features characteristic of individual faces.

21. The method according to claim 6, wherein said face regions include coefficients of the eigen-face representation characteristic of individual faces.

22. A method of generating an index of human faces appearing in a sequence of video frames, comprising:

processing said sequence of video frames to extract therefrom facial features; and

grouping said video frames to produce face tracks of contiguous frames, each face track being identified by the starting and ending frames in the track and containing face characteristic data of an individual face.

23. The method according to claim 22, wherein said face tracks are stored in a face index for use in searching or browsing for individual faces.

24. The method according to claim 22, wherein said face characteristic data includes eye, nose and mouth templates.

25. The method according to claim 22, wherein said face characteristic data includes image coordinates of geometric face features.

26. The method according to claim 22, wherein said face characteristic data includes coefficients of the eigen-face representation.

27. The method according to claim 22, wherein said grouping of video frames to produce said face tracks includes:

detecting predetermined facial features in a frame and utilizing said facial features for estimating the head boundary for the respective face;

opening a tracking window based on said estimated head boundary; and

utilizing said tracking window for tracking sequented frames which include said predetermined facial features.

28. The method according to claim 22, wherein said sequence of video frames is processed to also include audio data associated with said face tracks.

29. The method according to claim 22, wherein said sequence of video frames is processed to also include annotations associated with said face tracks.

30. The method according to claim 22, wherein said sequence of video frames is processing to also include face characteristic views associated with said face tracks.

31. The method according to claim 22, wherein said grouping includes merging tracks containing similar faces from a plurality of said face tracks.

32. A method of generating an index of at least one class of objects appearing in a collection of images to aid in browsing or searching for individual members of said class, comprising:

processing said collection of images to generate an index of features characteristic of said class of objects; and

annotating said index with annotations.

33. The method according to claim 32, wherein said method further comprises grouping said images in said index according to features helpful in identifying individual members of said class of objects.

34. The method according to claim 33, wherein said collection of images are a sequence of video frames.

35. The method according to claim 34, wherein said grouping of images, according to features helpful in identifying individual members of said class, produces a track of sequential frames for each individual member of said class of objects to be identified.

36. The method according to claim 35, wherein a listing of said annotations is generated and displayed for at least some of said tracks.

37. The method according to claim 35, wherein said at least one class of objects are human faces, and said grouping of images produces face tracks of contiguous frames, each track being identified by the starting and ending frames and containing face data characteristic of an individual face.

38. The method according to claim 37, wherein said sequence of video frames is processed to also include audio data associated with said face tracks.

39. The method according to claim 37, wherein said sequence of video frames is processed to also include face characteristic views associated with said face tracks.

40. The method according to claim 32, wherein said annotations include applying descriptions to various entries in said index.

41. The method according to claim 40, wherein said descriptions are applied manually during an editing operation.

42. The method according to claim 40, wherein said descriptions are applied automatically by means of a stored dictionary.

43. A method of processing a sequence of video frames having a video track and an audio track, to generate a speech annotated face index associated with speakers, comprising:

generating a face index from the video track;

generating a transcription of the audio track; and

aligning said transcription with said face index.

44. The method according to claim 43, wherein said transcription is generated by closed-caption decoding.

45. The method according to claim 43, wherein said transcription is generated by speech recognition.

46. The method according to claim 43, wherein said alignment is done by:

identifying start and end points of speech segments in said transcription;

extracting from the face index start and end points of face segments; and

outputting all speech segments that have non-zero temporal intersection with the respective face segment.

47. A method of processing a video track having face segments to label such segments as talking or non-talking, comprising:

tracking said face segments in the video track;

detecting segments having mouth motion; and

estimating from said detected segments, those having talking mouth motion vs. non-talking mouth motion.

48. The method according to claim 47, wherein segments estimated to have talking mouth motion are enabled for attaching speech to a speaker in a face segment.

49. A method of annotating a sequence of video frames, comprising:

processing said sequence of video frames to generate a face index having a plurality of entries; and

attaching descriptions to entries in said face index.

50. The method according to claim 49, further comprising attaching descriptions from said face index to at least one video frame.

51. A method of processing a sequence of video frames having a video track and an audio track, comprising:

extracting from said video track, face segments representing human faces, and producing a face track for each individual face;

extracting audio segments from said audio track;

fitting a model based on a set of said audio segments corresponding to the individual face of a face track; and

associating said model with the face track of the corresponding individual face.

52. The method according to claim 51, wherein said audio segments are speech segments.

53. Apparatus for generating an index of a class of objects appearing in a collection of images for use in searching or browsing for particular members of said class, comprising:

a processor for processing said collection of images to extract therefrom features characteristic of said objects, and for outputting therefrom indexing data with respect to said features;

a user interface for selecting the features to be extracted to enable searching for and identifying individual members of said class of objects; and

a store for storing said indexing data outputted from said processor in groups according to the features selected for extraction.

54. The apparatus according to claim 53, further including a browser-searcher for browsing and searching said store of indexing data to locate particular members of said class of objects.

55. The apparatus according to claim 54, further including an editor for scanning the indexing data store and for correcting errors occurring therein.

56. The apparatus according to claim 55, wherein said editor also enables annotating said indexing data store with annotations.

57. The apparatus according to claim 53, wherein said user interface enables selecting human facial features to be extracted from said collection of images to enable searching for individual human faces.

58. A method of searching a collection of images for individual members of a class of objects, comprising:

processing said collection of images to generate an index of said class of objects;

and searching said index for individual members of said class of objects.

59. The method according to claim 58, wherein said collection of images are a sequence of video frames.

60. The method according to claim 59, wherein processing said sequence of video frames produces a track of sequential frames for each individual member of said class of objects.

61. The method according to claim 60, wherein said class of objects are human faces.